What is claimed is:

1. A method for performing similarity searching by remote scoring and aggregating, comprising the steps of:

receiving a request by a similarity search server from one or more clients for initiating a similarity search, the request designating an anchor document and at least one search document;

generating one or more SQL commands from the client request;

sending the SQL commands from the similarity search server to one or more remote database management systems;

executing the SQL commands in the database management systems to determine normalized document similarity scores using user defined functions;

returning document similarity scores to the similarity search server from the one or more database management systems; and

constructing a search result and sending the search result to the one or more clients.

2. The method of claim 1, wherein the step of receiving a request from one or more clients further comprises generating one or more query commands for identifying a schema document for defining structure of search terms, mapping of datasets providing target search values to relational database locations, and designating measure, choice and weight algorithms to be used in a similarity search computation.

3. The method of claim 1, wherein the step of executing SQL commands further comprises using user defined functions contained within libraries of the database management systems for implementing measure algorithms to determine attribute similarity scores, weighting functions

and the choice algorithms for determining normalized document similarity scores, the document similarity scores being returned to the similarity search server.

4. The method of claim 1, wherein the step of executing the SQL commands further comprises:

computing attribute token similarity scores having values of between 0.00 and 1.00 for the corresponding leaf nodes of the anchor document and a search document using designated measure algorithms;

multiplying each token similarity score by a designated weighting function; and

aggregating the token similarity scores using designated choice algorithms for determining a document similarity score having a normalized value of between 0.00 and 1.00 for the search document.

5. The method of claim 2, wherein the step of generating one or more query commands comprises:

populating an anchor document with search criteria values;

identifying documents to be searched;

defining semantics for overriding parameters specified in an associated schema document;

defining a structure to be used by the result dataset; and

imposing restrictions on the result dataset.

6. The method of claim 5, wherein the step of defining semantics comprises:

designating overriding measures for determining attribute token similarity scores;

designating overriding choice algorithms for aggregating token similarity scores into document similarity scores; and

designating overriding weights to be applied to token similarity scores.

7. The method of claim 1 wherein the step of executing the SQL commands further comprises structuring the normalized similarity scores by imposing restrictions on the similarity scores according to a designated user defined function and returning restricted results to the similarity search server.

8. The method of claim 7, wherein the step of imposing restrictions is selected from the group consisting of defining a range of similarity scores to be selected and defining a range of percentiles of similarity scores to be selected.

9. The method of claim 1 wherein the step of executing the SQL commands further comprises sorting the normalized similarity scores according to a designated user defined function and returning sorted results to the similarity search server.

10. The method of claim 1 wherein the step of executing the SQL commands further comprises grouping the normalized similarity scores according to a designated user defined function and returning grouped results to the similarity search server.

11. The method of claim 1 wherein the step of executing the SQL commands further comprises executing statistics commands according to a designated user defined function and returning statistic results to the similarity search server.

12. The method of claim 1, wherein the step of executing the SQL commands to determine document similarity scores further comprises computing a normalized similarity score having a value of between 0.00 and 1.00, whereby a normalized similarity indicia value of 0.00 represents no similarity matching, a value of 1.00 represents exact similarity matching, and values between 0.00 and 1.00 represent degrees of similarity matching.

13. The method of claim 1, further comprising the steps of:

receiving a schema instruction from the one or more clients;

generating a schema command document including the steps of:

defining a structure of target search terms in one or more search documents;

creating a mapping of database record locations to the target search terms;

listing semantic elements for defining measures, weights and choices to be used in similarity searches; and

storing the schema command document into a database management system.

14. The method of claim 1, wherein the step of constructing and sending the search result to the one or more clients further comprises sending a response selected from the group consisting of an error message and a warning message.

15. The method of claim 1, wherein the step of constructing and sending the search result to the one or more clients further comprises sending a response to the one or more clients containing the result datasets, each result dataset including at least one normalized document similarity score, at least one search document name, a path to the search documents having a returned score, and at least one designated schema.

16. The method of claim 1, wherein the step of executing the SQL commands in the database management systems comprises executing one coalesced SQL search command to generate all similarity scores of multiple search documents for maximizing the processing once records have been loaded into memory and minimizing the number of disk accesses required.

17. The method of claim 1, wherein the step of executing the SQL commands comprises executing SQL commands in multiple database management systems for increased performance, each database management system containing a partition of a total target database to be searched.

18. The method of claim 17, further comprising the step of horizontally partitioning the total target database to be searched among the multiple database management systems.

19. The method of claim 17, further comprising the step of vertically partitioning the total target database to be searched among the multiple database management systems.

20. The method of claim 17, further comprising the step of horizontally and vertically partitioning the total target database to be searched among the multiple database management systems.

21. The method of claim 1, further comprising the step of selecting user defined functions for measure algorithms from the group consisting of name equivalents, foreign name equivalents, textual, sound coding, string difference, numeric, numbered difference, ranges, numeric combinations, range combinations, fuzzy, date oriented, date to range, date difference, and date combination.

22. The method of claim 1, further comprising the step of selecting user defined functions for choice algorithms from the group consisting of single best, greedy sum, overall sum, greedy minimum, overall minimum, and overall maximum.

23. A computer-readable medium containing instructions for controlling a computer system to implement the method of claim 1.

24. A system for performing similarity searching by remote scoring and aggregating, comprising:

means for receiving a request by a similarity search server from one or more clients for initiating a similarity search, the request designating an anchor document and at least one search document;

means for generating one or more SQL commands from the client request;

means for sending the SQL commands from the similarity search server to one or more remote database management systems;

means for executing the SQL commands in the database management systems to determine normalized document similarity scores using user defined functions;

means for returning document similarity scores to the similarity search server from the one or more database management systems; and

means for constructing a search result and sending the search result to the one or more clients.

25. The system of claim 24, wherein the means for receiving a request by a similarity search server is a gateway connected to a client network, the gateway also connecting to a search manager and a virtual document manager.

26. The system of claim 24, wherein the means for generating one or more SQL commands by the similarity search server is a search manager connected between a gateway and a database network interface.

27. The system of claim 24, wherein the means for sending the SQL commands from the similarity search server to one or more database management systems is a database network interface connected to a database network, the database network connecting to the database management systems.

28. The system of claim 24, wherein the means for executing the SQL commands is the database management systems, the database management systems including a library of user defined functions.

29. The system of claim 24, wherein the means for returning document similarity scores is the database management systems connected to a database network, the database network connecting to a database network interface of the similarity search server.

30. The system of claim 24, wherein the means for constructing a search result is a search manager and a virtual document manager within the similarity search server.

31. The system of claim 24, wherein the means for sending the search result to the client is a gateway connected to a search manager and a virtual document manager within the similarity search server, the gateway connecting to the one or more clients via a client network.

32. The system of claim 24, wherein the user defined functions are contained within libraries of the database management systems for implementing measure algorithms to determine attribute similarity scores, weighting functions and choice algorithms to determine normalized document similarity scores, the document similarity scores being returned to the similarity search server.

33. The system of claim 24, wherein user defined functions are contained within libraries of the database management systems for imposing restrictions on normalized similarity scores, sorting normalized similarity scores and grouping normalized similarity scores, the normalized similarity scores being returned to the similarity search server.

34. The system of claim 33, wherein the imposition of restrictions is selected from the group consisting of definition of a range of normalized similarity scores to be returned to the similarity search server and definition of a range of percentiles of similarity scores to be returned to the similarity search server.

35. The system of claim 24, wherein user defined functions are contained within libraries of the database management systems for determining statistics according to a designated user defined function, the statistic results being returned to the similarity search server.

36. The system of claim 32, wherein the normalized document similarity scores comprise a value of between 0.00 and 1.00, a normalized similarity value of 0.00 representing no similarity

matching, a normalized similarity value of 1.00 representing exact similarity matching, and normalized similarity values between 0.00 and 1.00 representing degrees of similarity matching.

37. The system of claim 24, wherein the request from one or more clients for initiating a similarity search comprises one or more query commands for identifying a schema document for defining structure of search terms, datasets for providing mapping of target search values to relational database locations, and designated measure, choice and weight algorithms to be used in a similarity search computation.

38. The system of claim 24, further comprising:

a gateway for receiving a schema instruction from the one or more clients;

a virtual document manager for generating a schema command document, including:

a structure of target search terms in one or more search documents;

a mapping of database record locations to the target search terms;

semantic elements for defining measures, weights and choices to be used in similarity searches; and

a database management for storing the schema command document.

39. The system of claim 24, wherein the means for constructing and sending the search result to the one or more clients further comprises means for sending a response selected from the group consisting of an error message and a warning message.

40. The system of claim 24, wherein the means for constructing and sending the search result to the one or more clients further comprises a gateway for sending a response to the one or more clients containing the result datasets, each result dataset including at least one normalized document similarity score, at least one search document name, a path to the search documents having a returned score, and at least one designated schema.

41. The system of claim 24, wherein the means for executing the SQL commands in the database management systems comprises one coalesced SQL search command to generate all similarity scores of multiple search documents for maximizing the processing once records have been loaded into memory and minimizing the number of disk accesses required.

42. The system of claim 24, wherein the means for executing the SQL commands comprises means for executing SQL commands in multiple database management systems for increased performance, each database management system containing a partition of a total target database to be searched.

43. The system of claim 42, further comprising means for horizontally partitioning the total target database to be searched among the multiple database management systems.

44. The system of claim 42, further comprising means for vertically partitioning the total target database to be searched among the multiple database management systems.

45. The system of claim 42, further comprising means for horizontally and vertically partitioning the total target database to be searched among the multiple database management systems.

46. The system of claim 24, further comprising user defined functions for measure algorithms selected from the group consisting of name equivalents, foreign name equivalents, textual, sound coding, string difference, numeric, numbered difference, ranges, numeric combinations, range combinations, fuzzy, date oriented, date to range, date difference, and date combination.

47. The system of claim 24, further comprising user defined functions for choice algorithms selected from the group consisting of single best, greedy sum, overall sum, greedy minimum, overall minimum, and overall maximum.

48. The system of claim 24, wherein:

the means for receiving a request by a similarity search server from one or more clients is via a secure client network connection; and

the means for sending the search result to the one or more clients is via a secure client network connection.

49. The system of claim 24, wherein:

the means for sending the SQL commands from the similarity search server to one or more remote database management systems is via a secure database network connection; and

the means for returning document similarity scores to the similarity search server from the one or more database management systems is via a secure database network connection.

50. A system for performing similarity searching by remote scoring and aggregating, comprising:

one or more clients communicating with a similarity search server for requesting a similarity search between an anchor document and at least one search document;

the similarity search server processing the similarity search request and constructing one or more SQL commands from the similarity search request;

the similarity search server communicating with one or more database management systems for transmitting the one or more SQL commands;

the one or more database management systems executing the SQL commands to obtain a similarity search result between the anchor document and the at least one search document;

the one or more database management systems communicating with the similarity search server for transmitting the search result; and

the similarity search server processing the similarity search result and communicating with the one or more clients for transmitting a similarity search response to the one or more clients.

51. The system of claim 50, further comprising a secure client network connection for transmitting a similarity search request and similarity search response between the one or more clients and the similarity search server.

52. The system of claim 50, further comprising a secure database network connection for transmitting the one or more SQL commands and the search results between the one or more database management systems and the similarity search server.